Contents

Problem 1

```
clear
close all

edges = csvread('wisconsin_edges.csv');

node_count = max(edges(:))+1;

A = zeros(node_count,node_count);
[m,n] = size(edges);
for i=1:m
  from_node = edges(i,1);
  to_node = edges(i,2);
  A(to_node+1,from_node+1)=1;
end

A = A+0.001;
A = A./sum(A);
[vecs,vals]=eig(A);
vals = diag(vals);
[val,ind] = max(vals);
index = eigs(A,1);
vec = vecs(:,ind);
S = A*vec;
for loop=1:100
    S=A*S;
end
[aa,indices]=sort(S,'descend');
first_index = indices(1,1)
third_index = indices(3,1)

% b) The most important page is Dane County, Wisconsin
% c) The third most important page is Sauk County, Wisconsin



% Hint: use
% eigs(A,k)
% where k=1 to get the first eigenvector, instead of
% eig(A)
% as computation of all eigenvectors will take ~5 minutes
```

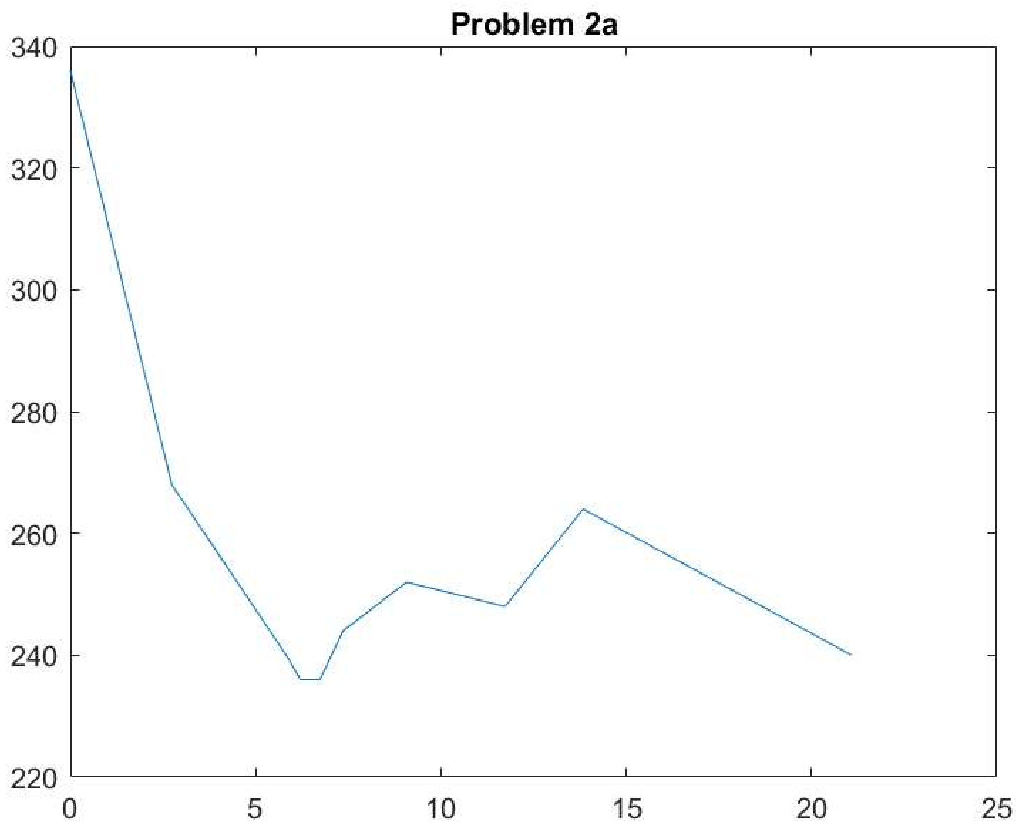
```
first_index =

        5090


third_index =
```

1346

Problem 2

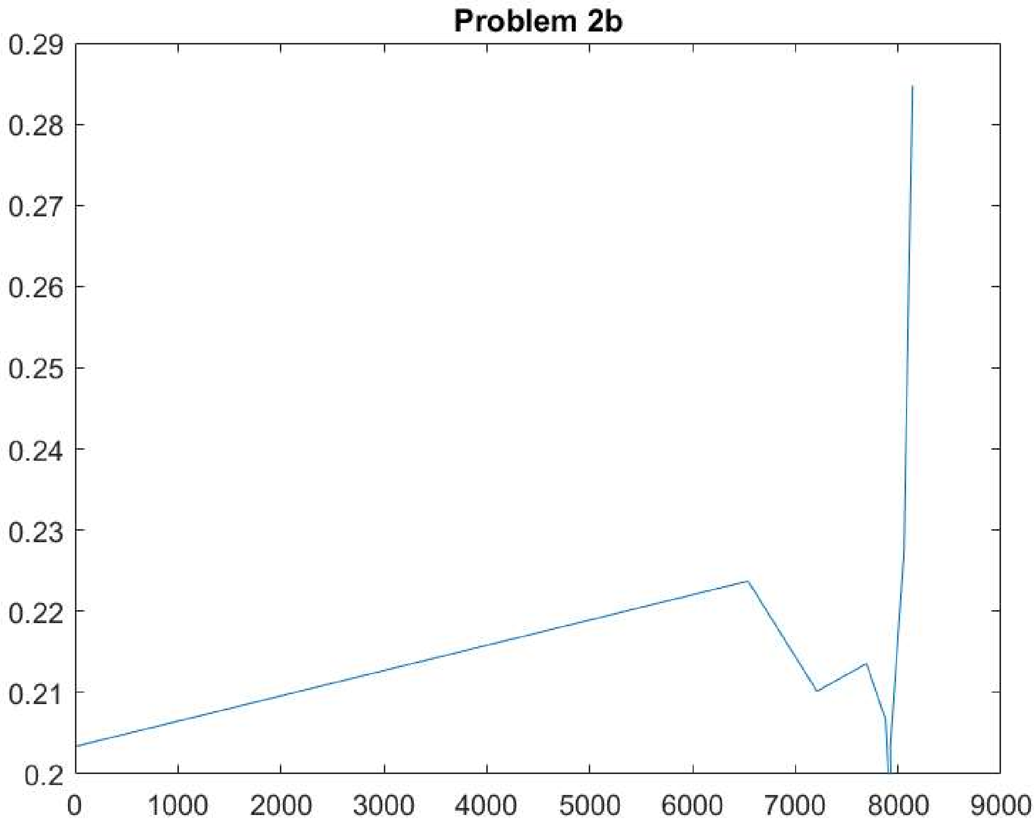
For the interpretation of each question in problem2, I included them as comments in matlab codes. Please check all of them. Thanks.

```
A = load('BreastCancer.mat');
x = A.X;
y = A.y;
vertical = [];
horizontal = [];
x1 = x(1:100,:);
y1 = y(1:100,1);
lambdas = [50,10,4.5,4,3.5,3,2,1,0.5,0];

weights = ista_solve_hot(x1,y1,lambdas);

for j=1:10
    lambda = lambdas(j);
    vertical(:,j) = norm(sign(x*weights(:,j))-y)^2;
    horizontal(:,j) = norm(weights(:,j),1);
end
plot(horizontal,vertical);
title('Problem 2a');

% According to the plot, the optimal weights go down when lambda increases.
% The plot shows that it becomes closer to zero in horizontal axis. The
% lowest point is the best solution. At the case when lambda is very small,
% weights will go up so that the error is larger.
```



Problem 2b

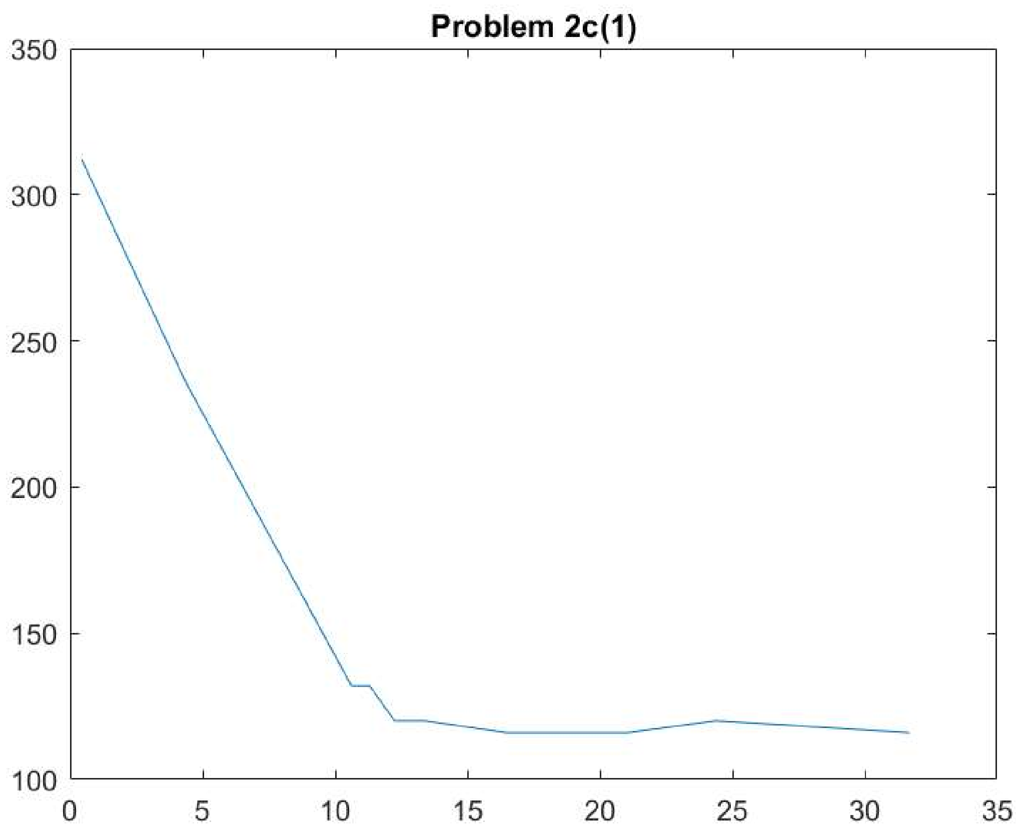
```
weights1 = weights;
weights1(abs(weights1) <= 10^(-6)) = 0;
sparsity = zeros(10,1);
error = zeros(10,1);

for j=1:10
    sparsity(j,1) = sum(weights1(:,j)==0);
    error(j,1) = sum(sign(x*weights(:,j))~=y)/295;
end
plot(sparsity,error);
title('Problem 2b');
% The error becomes larger when the sparsity increases. The number of
% features are smaller. When there is no regulation, the error rate is the
% highest. Also, sparsity is zero when there are a lot of regulations.
```

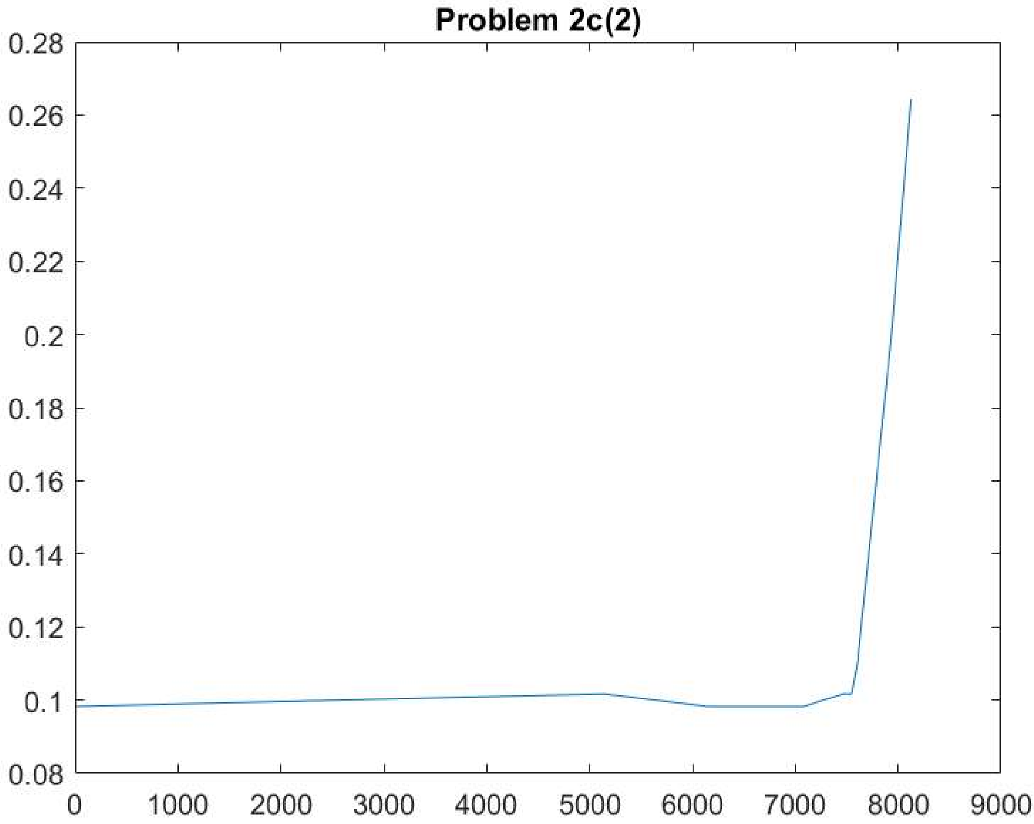


Problem 2c

```
vertical = [];
horizontal = [];
x1 = x(101:295,:);
y1 = y(101:295,1);
lambdas = [50,10,4.5,4,3.5,3,2,1,0.5,0];

weights = ista_solve_hot(x1,y1,lambdas);

for i=1:10
    lambda = lambdas(i);
    vertical(:,i) = norm(sign(x*weights(:,i))-y)^2;
    horizontal(:,i) = norm(weights(:,i),1);
end
figure;
```

```
plot(horizontal,vertical);
title('Problem 2c(1)');
% Optimal weights will go down when lambda increases. In particular, it's
% closer to zero in horizontal axis in the graph. The lowest point is the
% optimal. When lambda becomes small, optimal weights are almost the same
% (not change so much).

weights1 = weights;
weights1(abs(weights1) <= 10^(-6)) = 0;
sparsity = zeros(10,1);
error = zeros(10,1);

for j=1:10
    sparsity(j,1) = sum(weights1(:,j)==0);
    error(j,1) = sum(sign(x*weights(:,j))~=y)/295;
end
figure;
plot(sparsity,error);
title('Problem 2c(2)');
% When regulation exists, the error does not depend on sparsity or
% regulation very much. However, when there is lack of regulation, sparsity
% is large, and the error rate is large as well.
```





Published with MATLAB® R2016b

Breast Cancer LASSO Exploration

Contents

Prepare workspace

```
close all
clear

load BreastCancer
```

10-fold CV

```
% each row of setindices denotes the starting an ending index for one
% partition of the data: 5 sets of 30 samples and 5 sets of 29 samples
setindices = [1,30;31,60;61,90;91,120;121,150;151,179;180,208;209,237;238,266;267,295];

% each row of holdoutindices denotes the partitions that are held out from
% the training set
holdoutindices = [1,2;2,3;3,4;4,5;5,6;7,8;9,10;10,1];

cases = size(holdoutindices,1);

% be sure to initiate the quantities you want to measure before looping
% through the various training, validation, and test partitions
%
lam_vals = [1e-6 1e-4 1e-2 1e-1 logspace(0,2,20)];
squareerror_lasso = zeros(1,cases);
error_lasso = zeros(1,cases);
squareerror_rr = zeros(1,cases);
error_rr = zeros(1,cases);

% Loop over various cases
for j = 1:cases
    disp('Cases: ')
    j
    % row indices of first validation set
    v1_ind = setindices(holdoutindices(j,1),1):setindices(holdoutindices(j,1),2);

    % row indices of second validation set
    v2_ind = setindices(holdoutindices(j,2),1):setindices(holdoutindices(j,2),2);

    % row indices of training set
    trn_ind = setdiff(1:295,[v1_ind, v2_ind]);

    % define matrix of features and labels corresponding to first
    % validation set
    Av1 = X(v1_ind,:);
    bv1 = y(v1_ind);

    % define matrix of features and labels corresponding to second
    % validation set
    Av2 = X(v2_ind,:);
    bv2 = y(v2_ind);
```

```
    % define matrix of features and labels corresponding to the
    % training set
    At = X(trn_ind,:);
    bt = y(trn_ind);

% Use training data to learn classifier
    W = ista_solve_hot(At,bt,lam_vals);
    [m_w,n_w] = size(W);
    soln = zeros(1,n_w);
    for i=1:n_w
        soln(1,i) = norm(sign(Av1*W(:,i))-bv1)^2 + lam_vals(i)*norm(W(:,i),1);
    end

    [val,ind] = min(soln);
    final_result = sign(Av2*W(:,ind));
    squareerror_lasso(j) = norm(final_result-bv2)^2;
    error_lasso(j) = sum(final_result~=bv2);

    [f,g] = size(bv2);
    display('with LASSO: ')
    display('prediction error with each w is:')
    soln
    display('squared error on final subset with best lambda is: ')
    squareerror_lasso(j)
    display('test error on final subset with best lambda is: ')
    error_lasso(j)
    display('error rate: ')
    error_lasso(j)/f

    % ridge regression
    soln = zeros(1,n_w);
    M = transpose(At)*bt;
    N = transpose(At)*At;

    for k=1:n_w
        lambda = lam_vals(k);
        W = (N+lambda*eye(m_w))\M;
        soln(1,k) = norm(sign(Av1*W)-bv1)^2+lambda*norm(W)^2;
        weights(:,k)=W;
    end
    [val,ind]=min(soln);
    final_result = sign(Av2*weights(:,ind));
    squareerror_rr(j) = norm(final_result-bv2)^2;
    error_rr(j)=sum(final_result~=bv2);

    display('with ridge regression: ')
    display('prediction error with each w is:')
    soln
    display('squared error on final subset with best lambda is: ')
    squareerror_rr(j)
    display('test error on final subset with best lambda is: ')
    error_rr(j)
    display('error rate: ')
    error_rr(j)/f
end

mean_squareerror_rr = mean(squareerror_rr)
mean_error_rr = mean(error_rr)
mean_squareerror_lasso = mean(squareerror_lasso)
mean_error_lasso = mean(error_lasso)
```

```
Cases:

j =

     1

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   32.0001   32.0057   32.5704   37.7052   70.0191   75.3724   81.0831

  Columns 8 through 14

   87.3625   89.7863   92.2189   96.9277  101.3117  103.9669  103.0383

  Columns 15 through 21

   98.7542   94.9037   87.5971   84.2747   78.5731   73.8214   70.0676

  Columns 22 through 24

   70.5176   54.2212   45.7515

squared error on final subset with best lambda is:

ans =

   52.0000

test error on final subset with best lambda is:

ans =

    13

error rate:

ans =

    0.4333

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   32.0000   32.0001   32.0087   32.0866   32.8536   33.0828   33.3720

  Columns 8 through 14

   33.7357   34.1918   34.7612   35.4682   36.3406   37.4084   38.7026

  Columns 15 through 21
```

```
   40.2529   42.0839   44.2102   46.6312   49.3247   56.2424   59.3066

  Columns 22 through 24

   58.4119   57.4302   60.2223

squared error on final subset with best lambda is:

ans =

   52.0000

test error on final subset with best lambda is:

ans =

    13

error rate:

ans =

    0.4333

Cases:

j =

     2

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   48.0001   48.0058   48.5761   53.7622   83.9019   89.3993   95.3533

  Columns 8 through 14

  101.6581  108.1985  118.0240  127.1144  127.6762  131.6158  131.1384

  Columns 15 through 21

  117.7007  107.5965   97.9249   81.5413   89.0272   83.4812   68.3732

  Columns 22 through 24

   58.0338   55.4182   44.0528

squared error on final subset with best lambda is:

ans =

   20.0000

test error on final subset with best lambda is:

ans =
```

```
     5

error rate:

ans =

    0.1667

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   48.0000   48.0001   48.0089   48.0892   48.8796   49.1160   49.4144

  Columns 8 through 14

   49.7899   50.2609   50.8495   51.5811   52.4846   53.5920   54.9360

  Columns 15 through 21

   56.5484   58.4556   60.6735   63.2017   66.0162   69.0639   72.2592

  Columns 22 through 24

   75.4850   78.5993   81.4474

squared error on final subset with best lambda is:

ans =

   28.0000

test error on final subset with best lambda is:

ans =

     7

error rate:

ans =

    0.2333

Cases:

j =

     3

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7
```

```
   16.0001   16.0057   16.5748   21.7487   50.5960   55.9075   61.7520

  Columns 8 through 14

   67.5959   73.9319   80.3948   81.6982   85.3099   95.1796   99.2653

  Columns 15 through 21

   96.8848   91.5687   80.8517   73.9234   66.7481   67.5573   61.1651

  Columns 22 through 24

   55.6790   53.3469   56.2745

squared error on final subset with best lambda is:

ans =

   52.0000

test error on final subset with best lambda is:

ans =

    13

error rate:

ans =

    0.4333

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   16.0000   16.0001   16.0092   16.0919   16.9040   17.1463   17.4515

  Columns 8 through 14

   17.8350   18.3151   18.9132   19.6543   20.5660   21.6781   23.0207

  Columns 15 through 21

   24.6215   26.5025   28.6748   31.1337   33.8531   36.7816   39.8398

  Columns 22 through 24

   42.9220   45.9012   48.6390

squared error on final subset with best lambda is:

ans =

   52.0000

test error on final subset with best lambda is:
```

```
ans =

    13

error rate:

ans =

    0.4333

Cases:

j =

     4

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   56.0001   56.0058   56.5828   61.8286   94.0576   99.2617  105.1197

  Columns 8 through 14

  111.3162  117.9164  123.8539  129.5906  129.7320  128.3235  126.6415

  Columns 15 through 21

  120.2455  117.0867  109.2208  100.6102   99.8769   98.7062   91.9348

  Columns 22 through 24

   87.0395   76.8831   74.7793

squared error on final subset with best lambda is:

ans =

   20.0000

test error on final subset with best lambda is:

ans =

     5

error rate:

ans =

    0.1667

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7
```

```
   56.0000   56.0001   56.0093   56.0930   56.9148   57.1600   57.4692

  Columns 8 through 14

   57.8576   58.3440   58.9503   59.7017   60.6266   61.7554   63.1188

  Columns 15 through 21

   64.7452   66.6567   68.8644   71.3627   74.1235   81.0929   84.1883

  Columns 22 through 24

   87.3014   90.3039   93.0582

squared error on final subset with best lambda is:

ans =

   24.0000

test error on final subset with best lambda is:

ans =

     6

error rate:

ans =

    0.2000

Cases:

j =

     5

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   20.0001   20.0061   20.6121   26.1223   48.2000   54.1355   55.7909

  Columns 8 through 14

   61.8487   69.0032   78.6124   83.7409   87.1941   89.4797   89.1427

  Columns 15 through 21

   94.1687   84.3818   74.9673   71.8069   66.5971   61.5388   57.8019

  Columns 22 through 24

   52.2107   51.4001   35.4946

squared error on final subset with best lambda is:
```

```
ans =

    36

test error on final subset with best lambda is:

ans =

     9

error rate:

ans =

    0.3103

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   20.0000   20.0001   20.0102   20.1017   20.9995   21.2671   21.6041

  Columns 8 through 14

   22.0274   22.5568   23.2158   24.0313   25.0330   26.2525   27.7209

  Columns 15 through 21

   29.4659   31.5074   33.8519   36.4867   39.3744   42.4490   41.6153

  Columns 22 through 24

   44.7520   47.7208   50.3781

squared error on final subset with best lambda is:

ans =

    36

test error on final subset with best lambda is:

ans =

     9

error rate:

ans =

    0.3103

Cases:

j =

     6
```

```
with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   44.0001   44.0059   44.5863   49.8641   75.5406   76.9489   82.7176

  Columns 8 through 14

   89.1256   95.5085  101.8399  107.8491  114.2810  116.8585  118.4320

  Columns 15 through 21

  115.9740  100.6276   89.1169   82.6265   73.9693   59.4490   53.3394

  Columns 22 through 24

   54.6819   37.4185   21.9249

squared error on final subset with best lambda is:

ans =

   24.0000

test error on final subset with best lambda is:

ans =

     6

error rate:

ans =

    0.2069

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   44.0000   44.0001   44.0094   44.0936   44.9212   45.1681   45.4793

  Columns 8 through 14

   45.8703   46.3599   46.9700   47.7261   48.6567   49.7925   51.1645

  Columns 15 through 21

   52.8019   54.7281   56.9559   59.4829   62.2855   65.3151   68.4954

  Columns 22 through 24

   67.7228   70.8709   73.7986
```

```
squared error on final subset with best lambda is:

ans =

   20.0000

test error on final subset with best lambda is:

ans =

     5

error rate:

ans =

    0.1724

Cases:

j =

     7

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   44.0001   44.0056   44.5636   49.6360   74.9974   83.7350   88.6021

  Columns 8 through 14

   94.1640   99.3207  105.1595  109.8442  114.6878  117.5009  118.6997

  Columns 15 through 21

  117.2809  106.4847   97.2751   92.8386   89.7060   91.7862   83.5889

  Columns 22 through 24

   76.5800   69.7373   65.9121

squared error on final subset with best lambda is:

ans =

    56

test error on final subset with best lambda is:

ans =

    14

error rate:

ans =
```

```
    0.4828

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   44.0000   44.0001   44.0089   44.0892   44.8765   45.1110   45.4063

  Columns 8 through 14

   45.7769   46.2402   46.8165   47.5291   48.4036   49.4669   50.7455

  Columns 15 through 21

   52.2626   54.0344   56.0659   58.3454   60.8413   63.4987   66.2401

  Columns 22 through 24

   68.9691   71.5779   73.9579

squared error on final subset with best lambda is:

ans =

   52.0000

test error on final subset with best lambda is:

ans =

    13

error rate:

ans =

    0.4483

Cases:

j =

     8

with LASSO:
prediction error with each w is:

soln =

  Columns 1 through 7

   52.0001   52.0056   52.5653   57.6533   84.0250   89.1693   90.2288

  Columns 8 through 14

   99.6246  105.4405  111.4022  115.8792  119.5966  122.1775  122.1996

  Columns 15 through 21
```

```
  127.8698  124.9246  119.4569  111.4337  106.7382  106.6959  103.4403

  Columns 22 through 24

  100.8495   95.6109   93.0945

squared error on final subset with best lambda is:

ans =

   24.0000

test error on final subset with best lambda is:

ans =

     6

error rate:

ans =

    0.2000

with ridge regression:
prediction error with each w is:

soln =

  Columns 1 through 7

   48.0000   48.0001   48.0090   48.0900   48.8847   49.1214   49.4194

  Columns 8 through 14

   49.7934   50.2609   50.8426   51.5618   52.4443   53.5176   54.8084

  Columns 15 through 21

   56.3404   58.1305   60.1840   66.4905   69.0187   71.7145   74.5007

  Columns 22 through 24

   77.2804   79.9443   86.3808

squared error on final subset with best lambda is:

ans =

   24.0000

test error on final subset with best lambda is:

ans =

     6

error rate:

ans =
```

```
    0.2000


mean_squareerror_rr =

    36


mean_error_rr =

     9


mean_squareerror_lasso =

   35.5000


mean_error_lasso =

    8.8750
```

Published with MATLAB® R2016b